DOCUMENT RESUME

ED 039 226 TE 001 798

AUTHOR	Gibson, Susanne K.
TITLE	Computer-aided Construction of Categorized Spelling Tests.
INSTITUTION	Washington Univ., Seattle. Bureau of Testing.
PUB DATE	Jun 69
NOTE	25p.; Bureau of Testing Project 107, Research Summary
EDRS PRICE	EDRS Price MF-$0.25 HC-$1.35
DESCRIPTORS	*Computer Oriented Programs, *Methods Research, Multiple Choice Tests, Pattern Recognition, Programed Materials, *Spelling Instruction, *Test Construction, Testing Problems, Test Interpretation, Verbal Tests

ABSTRACT

To combat difficulties of dictated spelling tests, such as unreliable scoring due to illegible writing and the possibility of clues being provided through the enunciation of words by examiners, a technique was developed for writing experimental spelling tests by computer. Additionally, the diagnostic function of spelling scales was considered through the use of specific error categories in test construction. Five 50-item forms (each item containing from zero to four misspellings and representing a particular well-defined error category) were constructed and administered to 335 high school seniors, along with a standard battery of verbal tests. Although the machine-scorable tests raise their own difficulties--they may depend more on the student's proofreading ability than on his spelling proficiency and may, in the course of testing, teach the student misspelled words and so impair his spelling ability--the experimental tests were found to function as well as the standard spelling test. The 12 separate error categories did not function as independently as was hoped; the need for further experimentation with four "super" categories (addition, omission, inversion, and substitution) was indicated. (Twenty sample test items, tables of test results, and a list of references are included.) (Author/MF)

N-X

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

HIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE
ERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS
TATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION
OSITION OR POLICY.

Bureau of Testing

University of Washington

June 1969

Computer-aided Construction of Categorized Spelling Tests

Susanne K. Gibson

ED039226

A technique was developed for writing experimental spelling tests by computer. Five fifty-item forms were constructed, each item containing from zero to four misspellings and representing a particular well-defined error category. These tests were administered together with a standard battery of verbal tests. The experimental tests functioned as well as the standard spelling test. The usefulness of error categories was, however, questioned.

Although research in the field of spelling has had considerable scope recent spelling finds have not reached the classroom. So long as the spelling of English words remains conventional, rather than phonetic, and so long as misspellings continue to be regarded as both academically and socially symptomatic, there will be a need for a reliable and economic method of testing spelling performance--both for measurement of spelling ability as a subskill of composition and for research into spelling problems (Thomas, 1968). Investigation of the domain of spelling ability in relation to other areas of cognitive functioning has suggested that spelling is a relatively independent skill (Thurston, 1948; Knoell & Harris, 1952; Ager & Allen, 1965). In their recent factor analytic study of spelling ability, Ager and Allen (1965) established this independence from three variables previously considered critical to spelling ability; spatial ability, authoritarianism, and verbal reasoning.

ED039226

Bureau of Testing Project: 107. This report summarizes research done by the author as a candidate for the Honors B.S. in Psychology.

Agreement of two early studies (Wager, 1912: Tidyman & Johnson, 1924) indicates that grouping words by similarity and structure is of assistance in learning to spell such words. Foran's (1934) survey of the grouping problem supported the belief that grouping should be practiced except in specific, disadvantageous cases, such as homonyms. Furthermore, if spelling mistakes can be classified into certain types, these are grounds for teaching words in groups in order to guard against mistakes made persistently (Foran, 1934). Numerous studies of spelling errors have shown that mistakes should be differentiated into categories of errors, which most writers define as classes in which mistakes may be grouped (Hollingworth, 1919; Book & Harter, 1926; Masters, 1927; Alper, 1942). Alper has further differentiated error categories into objective and subjective categories. She considers objective categories, such as omission of letters, failure to double a consonant, etc., easy to set up but limited in value, as opposed to subjective categories. Such subjective categories including tension at a hard spot of a word, as in the case of reversals or anticipated letter insertions, and negative transfer, as in the case of incorrect carry-over of the spelling of the root or base of a word, are difficult to set up, because operational definitions are difficult to formulate (Alper, 1942). Most of the error analysis studies attempt to assert that errors represent not so much an inability to spell some words as an indifferent attitude toward mistakes, themselves (Foran, 1934).

High correlations between various spelling measures which don't differ in factorial composition suggest that spelling ability is not influenced by the method of testing (Ager & Allen, 1965). However, most results in

experimental reports over many years show that although the scores on recognition and recall spelling tests often yield intercorrelations close to .90, the recognition scores are higher than the recall scores and the scores on particular words vary widely between the two tests according to the difficulty of the words and maturity of the students (Lindquist & Cook, 1933; Weller & Broom, 1934; Nelson & Denny, 1936; Moore, 1937, Sturdyvin, 1937; Jackson, 1943; Brody, 1944). Nevertheless, scoring tedium and expense of recall tests represented by the traditional dictation forms have led to wide use of machine-scorable tests which depend on student recognition of correctly spelled or misspelled words. Specific faults of the dictated tests include: (1) unreliable scoring due to illegible writing, (2) presence of clues to correct spelling in the enunciation of problem words by the examiner, (3) difficulty of transferring item data from the test by hand to machine-scorable answer sheets or computer cards before statistical analysis of test outcomes can be undertaken, and (4) known deterioration of the spelling ability of the scorer himself over long periods of scanning misspelled words. However, there are also objections to machine-scorable tests: (1) The method lacks natural relevance, since the student's performance is not an act of spelling and must depend to some extent on visual acuity or proofreading ability, rather than spelling proficiency; (2) the student may learn, in the course of testing, misspelled words and so impair his spelling ability; (3) above all, the test does not yield the same difficulty coefficients for the same words when they are presented in misspelled or correctly spelled forms, i.e., students are more likely to accept the misspelling of a word than to reject its correct spelling (Wesman, 1946; Thomas, 1968). In a comparison of two tests in which half of the words correctly spelled in one test were incorrectly spelled in

the other test and the other half which were incorrectly spelled in the former test were correctly spelled in the latter, Wesman (1946) found a difference in validity of correct and incorrect forms. A comparison of the coefficients of correlation for each item in its correct and incorrect forms against the total score yielded higher coefficients with the incorrect forms. In light of indications that only the incorrectly spelled words on a machine-scorable recognition test contribute to the testee's score, Wesman suggested that a fully efficient spelling test would be of the true-false type in which the testee must consider both correct and incorrect words. Wesman, however, did find that a certain number of words were useful when they were correctly spelled (Wesman, 1946).

Not only can tests be scored rapidly and reliably by machine, a recent study demonstrated the usefulness of a computer in writing items (Fremer & Anastasio, 1968). Spelling items were used, because they seemed to have the fewest facets or dimensions. This study developed and programmed rules for item construction which were then applied to selected words. The item construction program was used with a set of 40 rules for creating misspelled words. In their evaluation of the computer output, experienced spelling test item writers concluded that the computer-generated lists of words with their misspellings would be a helpful resource even with a large proportion of implausible misspellings. Woodsen (1968) who recognized the problem of selecting and ordering items in test construction also provided a computer-programmed solution. The program input included the set of n items and their answers; the output included any number of items randomly selected from the set of n items and randomly ordered.

The present study was undertaken to respond to several disadvantages of current spelling tests by developing a machine scorable test in which randomly ordered items can be answered in a true-false or forced-choice form. Items for such a test should be based on randomly selected words representing a wider range of critical words than contribute to present machine-scorable tests. Additional goals of the study were to consider the diagnostic function of spelling scales through the use of specific error categories in test construction and to investigate whether grouping of words, similarly misspelled, into items representing distinct error categories would facilitate discrimination of the correctness or incorrectness of a word's spelling.

Method

The words which constituted the item pool were drawn from several sources of frequently misspelled words. A list of 606 "most frequently misspelled words in the English language" (Furness, 1964) was used as a starting place. From this same source the entire "Remington List of Words Most Frequently Misspelled by Adults" was also drawn. Other words were selected from a list of words most frequently taught in U.S. classrooms (Gates, 1937). Words from these lists rated below 6.0 (sixth grade) difficulty were not employed. Additional words were then drawn from a list of 30,000 words for which numerical ratings of frequency of occurrence in general and in four different sets of reading material were known (Thorndike & Lorge, 1944). In this selection words were included which occurred at least once per one million words but less often than 15 times per million words. Many of the words included in this last selection were words whose roots were among the shorter

words which had been earlier excluded on the basis of their Gates gradings (below 6.0).

On separate 3x5 index cards each word along with one or more common misspellings, difficult letters underscored, number of letters in the correctly spelled form of the word, and ratings from various sources were recorded. After words from the first two sources were recorded, cards containing words rated below 6.0 or having more than 15 or fewer than 4 letters were separated as reference for selecting words from the sub-set of the 30,000 word list. After words were selected from the latter source, a second set of words was deleted from the item pool. These words included proper nouns, separated or hyphenated words (e.g. all right, go-between, etc.) and those words whose misspelled forms were actually correctly spelled forms of different words (i.e. homonyms).

In order to categorize the errors in each misspelled form of a word, a list of 20 different error categories was compiled from word structuring categories (Furness, 1964) and from Master's Classification of Spelling Errors (Foran, 1934).

Clerks in the Bureau of Testing were asked to assign each word to one of the 20 error categories. They specified the category by number and indicated the letter(s) involved in generating the misspelled form of the word. Because clerks encountered difficulty with the structuring categories, these categories were eliminated. Accordingly, the error categories actually used were:

(1) Double vowel for a single vowel
(2) Double consonant for a single consonant
(3) Single vowel for a double vowel
(4) Single consonant for a double consonant
(5) Omission of vowels
(6) Omission of consonants
(7) Addition of vowels
(8) Addition of consonants

(9) Inversion of letters (e.g. spelling "ei" instead of "ie")
(10) Phonic substitution of a vowel (or vowel pair) for another similar in sound
(11) Phonic substitution of one consonant (or consonant pair) for another similar in sound
(12) Phonic substitution of a vowel-consonant pair for another vowel-consonant pair similar in sound (including also vowel-pair substitutions for vowel-consonant pairs)

These words with their descriptive error classifications were then punched on cards. Each card included one correctly spelled word, one misspelled form of that word, and the appropriate error classification. This classification included both the error category number and a letter code which identified the letters in the correctly spelled form which were involved or distorted in the misspelled form. For the addition or insertion of letters category the letter code identified the additional letter. In order to determine the actual sets of difficult letters involved in each error category, cards were then sorted by their number-letter classification. Within each numerical error category, all words were sorted by letter classification, into alphabetical order and grouped into the letter sub-groups within each numbered category. Sub-categories with fewer than six entries were dropped at this stage. Category 1, doubling vowels, was dropped because of the few examples available. All sub-categories were required to contain a minimum nunber of words so that a variety of items could be constructed from each sub-category.

A computer program was then designed to construct tests from the remaining set of words and their misspelled forms according to the following specifications:

(1) Fifty-five items were to be selected for each test
(2) Each item was to have four words randomly drawn from the same error sub-category
(3) To construct a single item the program had to randomly select:
    (a) an error sub-category
    (b) four words from that error sub-category
    (c) either a correct or incorrect spelling of each of the four words.

Ten sets of fifty-five items were to be constructed in this way. If any of the first fifty items of a given set included word repetitions (random selection did not preclude selecting the same word more than once) they were to be dropped and one of the last five items substituted to produce a final form of each test containing fifty items. Tests, test directions and answer sheets were prepared. For these tests subjects were to be asked to indicate whether each of the four words making up each item was correct or incorrect. Figure 1 shows representative items drawn from Form D of these tests.

In order to determine the length of time necessary to complete such a test, a preliminary session examined 50 college freshman subjects, randomly drawn from Introductory Psychology classes. All Ss received one form of the ten tests in this session. On the answer sheets provided for the Ss, a large blank square appeared below each set of ten item answer blanks. As Ss completed each ten items, they looked at the clock and recorded the time in the blank square. Because E recorded the starting time, Ss only recorded the clock time after the tenth, twentieth, thirtieth, fortieth and fiftieth items. This testing session demonstrated that a fifteen minute time period would be sufficient for the average subject to complete the test.

During Fall quarter 1968, high school students taking the Washington Pre-College battery on the UW campus were tested using five forms of this experimental spelling test. Each of these 335 high school seniors received one form of the five as the last test of their morning battery of tests. The experimental form appeared on a separate mimeographed form similar to the Washington Pre-College Spelling section. Ss were allowed 15 minutes to complete the experimental test.

Figure 1

Test Items, Experimental Spelling Test, Form D

1. intimate, harmed, cucumnber, obstinate
2. suscriber, doubtful, dout, adominal
3. underwear, boulaverd, congragetion, elagence
4. haggard, strugle, sugestion, exagerate
5. bugget, consider, divident, graduation
6. deceive, recieving, sieze, bulliten
7. almanac, particulary, amighty, regularly
8. pronounciation, ancestral, coustom, abuse
9. ambalunce, gaurantee, gradually, circalur
10. faith, attainable, entertane, certenly
11. arrears, spere, weary, harbor
12. pospone, delightful, bankruptcy, fiction
13. inform, income, confirm, imformed
14. emmanate, ommit, demmand, humility
15. wiccid, gnit, remarkable, racket
16. agreement, feble, exceedingly, indeed
17. licquor, arctical, awkward, liquid
18. earge, searum, cafetearia, anaesthetic
19. principle, absorption, subscribtion, adobt
20. secretary, massacre, acre, mediocer

These tests were then scored two ways, by words and by items. Word scores represented total number of words correctly marked less the number of words incorrectly marked. Item scores represented the total number of items correct. To have an item correct all four words making up that item had to be correctly marked.

In order to score these tests according to error categories the ten categories had to be regrouped into a smaller number of error "super" categories. These error "super" categories and the regrouped error categories which they represented were:

I. Addition: Insertion of a letter within the correct spelling of a word
   - Cat. 2: Double consonant instead of a single consonant (Insertion of a consonant identical to an adjacent consonant essential to the correct spelling of the word)
   - Cat. 7: Non-doubling vowel additions (Insertion of a vowel which may or may not be identical to vowel(s) essential to the correct spelling)
   - Cat. 8: Non-doubling consonant additions (Insertion of a consonant which may or may not be identical to consonant(s) essential to the correct spelling)

II. Omission: Omission of a letter essential to the correct spelling of a word
   - Cat. 3: Single vowel for double vowel (Omission of a vowel from a pair of identical vowels)
   - Cat. 4: Single consonant for a double consonant (Omission of a consonant from a pair of identical consonants)
   - Cat. 5: Non-pair-splitting vowel omissions (Omission of a vowel or vowel pair, which function as a single vowel, from the correct spelling)
   - Cat. 6: Non-pair-splitting consonant omissions (Omission of a consonant or consonant pair, which function as a single consonant, from the correct spelling)

III. Inversions: Reversals of the order of two letters (including vowel pair or vowel/consonant combinations) which may or may not involve an arbitrary third letter; Cat. 9 defines this super category

IV. Substitutions: Replacement of one letter (or letter pair) essential to the correct spelling of the word with another letter (or letter pair)

Cat. 10: Vowel substitutions (replacement of one vowel or pair with another vowel or pair)

Cat. 11: Consonant substitutions (replacement of one consonant or pair with another consonant or pair)

Cat. 12: Vowel-consonant substitutions (replacement of a vowel-consonant pair with another vowel-consonant pair)

All further analyses and discussion depend on these "super" categories and the ten original error categories are not considered further. Each test was then scored by items and words for each of the categories. Scores for total items and words correct and for items and words correct for each category were compared within each experimental form. Total scores and category scores on each form were also compared with scores on four of the verbal tests included in the Washington Pre-College test battery.

Results

Mean scores for correct words and items, scored totally and for each category, are shown in Table 1. For the total scores the maximum possible scores are the same for all forms: 200 words or 50 items. For categories, however, these maxima vary from form to form and the number of items belonging to each category in each form have been indicated by parenthesized numbers in Table 1. The corresponding number of words is, of course, four times that number. In Table 1 means are followed by SD's.

To permit more direct comparison of these mean results, Table 2 reports the mean proportions of words and items answered correctly on each of the five forms. Because all of the words making up an item had to be correct before an item was regarded as answered correctly the proportion of items

TABLE 1

Mean Scores and Standard Deviations for Experimental Spelling Tests

	Form 1 (N = 68)		Form 2 (N = 71)		Form 3 (N = 66)		Form 4 (N = 65)		Form 5 (N = 65)	
Total words	118.015	34.56	110.056	35.73	110.061	38.32	114.046	37.14	115.892	45.19
items	21.294	8.72	19.746	8.90	19.606	9.68	20.524	9.74	19.723	9.99
Cat. 1: words	32.353	10.98	37.042	13.47	31.590	12.35	37.631	43.52	29.554	13.76
items	5.721	2.83	6.732	3.50	5.121	2.98	6.538	3.56	4.769	2.89
	(14)		(16)		(16)		(17)		(15)	
Cat, II: words	41.485	11.85	25.521	7.47	34.167	12.77	34.461	10.39	30.169	12.01
items	7.441	3.30	4.049	1.98	5.924	3.29	6.446	3.08	5.185	3.03
	(16)		(12)		(16)		(14)		(13)	
Cat. III: words	16.691	6.02	14.606	6.64	22.197	10.28	20.092	10.01	18.108	7.48
items	3.059	1.54	2.521	1.65	4.258	2.52	3.585	2.41	3.123	1.94
	(8)		(8)		(10)		(10)		(8)	
Cat. IV: words	27.750	10.17	33.141	11.77	21.803	7.34	21.831	6.99	35.600	13.99
items	5.324	2.53	5.958	3.06	4.409	2.14	3.954	1.95	6.646	3.39
	(12)		(14)		(8)		(9)		(14)	

TABLE 2

Mean Proportion of Words and Items Correct

	Form 1	Form 2	Form 3	Form 4	Form 5	Average
Total words	.590	.550	.550	.570	.579	.569
items	.425	.394	.392	.410	.394	.403
Cat. I: words	.578	.579	.494	.553	.492	.538
items	.409	.421	.320	.385	.318	.369
Cat. II: words	.648	.532	.534	.615	.580	.581
items	.465	.379	.370	.461	.398	.414
Cat. III: words	.522	.456	.555	.502	.566	.519
items	.382	.315	.426	.358	.390	.374
Cat. IV: words	.578	.592	.681	.606	.636	.618
items	.443	.426	.551	.439	.475	.466

answered correctly is always smaller than the proportion of words answered correctly. Although there was form to form variability, Category IV items (or words), involving letter substitutions, appeared to be the easiest and Category III items, reversals of letters in a letter pair, the hardest. For the two forms on which Category IV items were not the easiest (Forms 1 and 4) this place was held by Category II, omissions of a letter, and for the two forms, 3 and 5, on which Category III was not the hardest this position was occupied by Category I, letter insertions.

Although the five experimental spelling forms were distributed at random among the 335 high school seniors tested such randomization could not be expected to assure that the five resulting groups were matched on any relevant aptitudes or other attributes. Table 3 reports mean scores earned by the five groups on four of the verbal tests of the WPC battery. Scores are in a standard score system for which the statewide high school senior mean is 50 and the standard deviation 10. Table 3 suggests that the group assigned experimental form 3 were less verbally proficient and the group completing form 4 more skilled than the remaining Ss.

Correlations among total word scores, total item scores, and word and item scores for each category within each form are shown in Tables 4-8. Correlations between item and word scores are indicated by parentheses. The number of test items represented by each category is indicated in parentheses preceding the heading "items" for each category. Correlations between total scores and the several category scores were influenced, of course, by the number of items (or words) that category contributed to the total score. Thus, Category III, with relatively fewer items tended to correlate with the total scores at a lower level than the other categories.

TABLE 3

WPC Verbal Test Means and Standard Deviations

for Experimental Spelling Test Groups

WPC test	Subjects									
	Form 1 (N = 68)		Form 2 (N = 71)		Form 3 (N = 63)		Form 4 (N = 65)		Form 5 (N = 65)	
	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.
Reading Comprehension	46.500	11.09	47.747	12.06	44.894	10.46	49.892	11.20	46.385	11.14
Vocabulary	48.132	10.55	47.303	10.31	46.682	10.26	49.738	10.28	47.231	11.34
English Usage	45.294	8.24	45.028	8.72	44.258	8.33	47.015	7.68	44.957	8.58
Spelling	49.029	9.42	49.239	9.73	46.636	9.09	49.738	8.78	47.554	9.79

TABLE 4

Correlations between Total Scores and Category Scores (Form 1)

	Category I		Category II		Category III		Category IV	
	words	items	words	items	words	items	words	items
Total words	.904	(.795)	.935	(.898)	.690	(.626)	.920	(.836)
items	(.859)	.821	(.826)	.902	(.688)	.722	(.821)	.842
Cat. I: words (14)	---	(.921)	.802	(.787)	.534	(.543)	.765	(.703)
items		---	(.802)	.710	(.534)	.523	(.649)	.632
Cat. II: words (16)			---	(.902)	.539	(.511)	.842	(.743)
items				---	(.502)	.599	(.812)	.791
Cat. III: words (8)					---	(.850)	.546	(.536)
items						---	(.476)	.507
Cat. IV: words (12)							---	(.899)
items								---

TABLE 5

Correlations between Total Scores and Category Scores (Form 2)

	Category I		Category II		Category III		Category IV	
	words	items	words	items	words	items	words	items
Total words	.932	(.889)	.845	(.700)	.843	(.691)	.911	(.903)
items	(.882)	.941	(.835)	.786	(.819)	.759	(.818)	.907
Cat. I: words (16)	---	(.925)	.836	(.578)	.722	(.598)	.801	(.806)
items		---	(.732)	.668	(.709)	.638	(.739)	.810
Cat. II: words (12)			---	(.886)	.684	(.578)	.672	(.703)
items				---	(.586)	.519	(.511)	.588
Cat. III: words (8)					---	(.842)	.699	(.731)
items						---	(.520)	.600
Cat. IV: words (14)							---	(.912)
items								---

TABLE 6

Correlations between Total Scores and Category Scores (Form 3)

	Category I		Category II		Category III		Category IV	
	words	items	words	items	words	items	words	items
Total words	.867	(.814)	.920	(.846)	.904	(.867)	.890	(.762)
items	(.756)	.753	(.882)	.896	(.854)	.887	(.780)	.842
Cat. I: words	---	(.893)	.708	(.627)	.695	(.648)	.689	(.592)
(16) items		---	(.681)	.672	(.636)	.647	(.729)	.488
Cat. II: words			---	(.908)	.793	(.780)	.795	(.715)
(16) items				---	(.753)	.814	(.729)	.698
Cat. III: words					---	(.944)	.795	(.715)
(10) items						---	(.772)	.698
Cat. IV: words							---	(.816)
(8) items								---

TABLE 7

Correlations between Total Scores and Category Scores (Form 4)

	Category I		Category II		Category III		Category IV	
	words	items	words	items	words	items	words	items
Total words	.930	(.850)	.910	(.900)	.989	(.826)	.874	(.777)
items	(.882)	.901	(.872)	.925	(.862)	.845	(.846)	.824
Cat. I: words	---	(.904)	.765	(.774)	.770	(.697)	.774	(.670)
(17) items		---	(.708)	.754	(.706)	.689	(.695)	.645
Cat. II: words			---	(.916)	.771	(.712)	.764	(.698)
(14) items				---	(.787)	.723	(.794)	.757
Cat. III: words					---	(.926)	.709	(.636)
(10) items						---	(.656)	.595
Cat. IV: words							---	(.886)
(9) items								---

TABLE 8

Correlations between Total Scores and Category Scores (Form 5)

	Category I		Category II		Category III		Category IV	
	words	items	words	items	words	items	words	items
Total words	.894	(.778)	.888	(.772)	.824	(.713)	.914	(.831)
items	(.900)	.894	(.892)	.877	(.868)	.872	(.845)	.887
Cat. I: words (15)	---	(.906)	.828	(.716)	.817	(.756)	.851	(.744)
items		---	(.729)	.682	(.800)	.783	(.708)	.910
Cat. II: words (13)			---	(.920)	.779	(.710)	.799	(.765)
items				---	(.720)	.721	(.651)	.682
Cat. III: words (8)					---	(.869)	.759	(.723)
items						---	(.653)	.672
Cat. IV: words (14)							---	(.919)
items								---

Between category correlations showed no clear pattern. At the word level, correlations tended to range between .7 and .8. The one exception was provided by Category III on Form 1. This 8 item category correlated only in the .50's with the other categories on this form. At the item level there is even greater variability but some evidence that categories I, II or IV correlated stronger with each other than did these categories with Category III.

Tables 9-12 show correlations of experimental spelling scores with each of the four WPC verbal tests. Correlations of the experimental spelling tests with the WPC spelling measure make up Table 9. Total item scores on the five experimental forms correlated from .6 for Form 3 to above .8 for Forms 2 and 5 with the WPC measure, a fifty item, five choice test. Each WPC item consisted of four correctly spelled and one incorrectly spelled word with the Ss task to identify the misspelled word. The WPC spelling score, prior to standardization, was obtained by subtracting one-fourth of the incorrectly answered items from the number answered correctly. This is, of course, the common correction for guessing. Estimated reliability for the WPC test, based on odd-even correlation is .85 (WPC, 1968).

The two experimental forms, 2 and 5, most highly correlated with the WPC test were marked by having fewer Category II (letter omission) and more Category IV (letter substitution) items than the remaining experimental tests. Possibly more responsible for the higher total correlations, however, were the correlations of Category I (letter addition) items with the WPC scores.

Table 10 presents correlations of the spelling tests with the WPC vocabulary test. This latter test consists of 100 five choice antonym items with scores again adjusted for guessing. Reliability has been estimated

TABLE 9

Experimental Spelling Test Correlati ons with

WPC Spelling Test

(Decimal points omitted)

		Form 1	Form 2	Form 3	Form 4	Form 5
Total words		704	844	628	665	736
items		665	850	605	673	843
Cat. I:	words	671	803	553	588	783
	items	611	818	615	600	837
Cat. II:	words	624	737	580	615	725
	items	617	632	636	598	732
Cat. III:	words	512	749	558	649	752
	items	448	682	560	616	756
Cat. IV:	words	642	716	551	548	632
	items	637	756	428	542	671

TABLE 10

Experimental Spelling Test Correlations with

WPC Vocabulary Test

(Decimal points omitted)

		Form 1	Form 2	Form 3	Form 4	Form 5
Total words		618	549	586	550	633
items		570	536	524	543	698
Cat. I:	words	550	544	533	583	648
	items	495	545	518	539	651
Cat. II:	words	552	390	507	375	644
	items	544	369	484	411	603
Cat. III:	words	488	496	546	551	597
	items	406	445	522	509	528
Cat. IV:	words	577	499	534	452	643
	items	572	470	409	466	649
WPC Spelling		607	563	518	490	703

TABLE II

Experimental Spelling Test Correlations with

WPC English Usage Test

(Decimal points omitted)

		Form 1	Form 2	Form 3	Form 4	Form 5
Total words		575	667	558	654	665
items		553	667	506	654	764
Cat. I:	words	568	659	525	651	728
	items	505	648	505	586	745
Cat. II:	words	473	521	515	507	689
	items	479	493	496	558	668
Cat. III:	words	473	507	490	657	646
	items	398	460	476	609	657
Cat. IV:	words	521	616	468	523	604
	items	554	617	369	560	631
WPC Spelling		635	714	685	629	776

TABLE 12

Experimental Spelling Test Correlations with

WPC Reading Comprehension Test

(Decimal points omitted)

		Form 1	Form 2	Form 3	Form 4	Form 5
Total words		605	595	562	478	518
items		531	586	510	503	606
Cat. 1:	words	547	626	466	516	556
	items	455	613	445	481	632
Cat. II:	words	553	505	548	317	486
	items	512	472	523	388	443
Cat. III:	words	449	414	524	427	558
	items	374	370	503	422	527
Cat. IV:	words	560	523	484	471	520
	items	546	493	375	484	539
WPC Spelling		523	612	556	520	643

at .95 (WPC, 1968). Total item scores for the experimental tests correlated with vocabulary at essentially the same level as the WPC spelling test with the vocabulary measure. There was, incidentally, considerable variability in these between WPC measure correlations. Estimates of the vocabulary-spelling correlation ranged from below .5 for the group taking experimental form 4 to above .7 for the group writing experimental test 5.

This confounding of groups with forms was further illustrated by the correlations of the experimental spelling forms with the remaining two WPC verbal tests, English usage and reading comprehension, presented in Tables 11 and 12. Total item scores for experimental forms 2 and 5 were again more highly correlated with the WPC measures than the other forms. At the same time, however, it is clear that English usage and reading comprehension were also more highly correlated with the WPC spelling test for the groups taking those two experimental forms.

Discussion

Through computer programming, the present study developed a battery of machine scorable spelling tests which required testees to respond to each of the randomly selected words in the randomly ordered items by forced-choice (correct vs. incorrect) methods. This forced choice method required that Ss respond to each of four words in each of 50 items. The comparable 50-item WPC spelling section which uses a multiple-choice answering method, required that Ss respond to only one "wrong" word in each of four words in each item. Initial testing with experimental battery demonstrated that the average testee needed at least fifteen minutes to complete the fifty items, while most testees required only ten minutes to complete the WPC spelling

section. This greater length of time necessary for experimental forms indicated that Ss had to respond to a greater number of critical words in the experimental forms.

Each of the fifty items represented a particular error category and one set of one or two letters involved in the error. Because categories were represented by a variable number of items correlations involving the categories were difficult to interpret. Perhaps equal representation of each category would better indicate the predictive values for each category of total scores.

High correlations among the error categories for adding, subtracting, and substituting letters erroneously, suggest that these errors involve similar processes. However, low correlations between Category III and the other categories suggest that the process of inverting the order of letters may be dissimilar to other error processes. Low mean scores on Category III across forms indicate that this type of error may also be more difficult to discriminate. Possibly the S recognizes that the correct letters are present in the word, but he fails to recognize that these letters appear in the wrong order. Consistently higher mean scores on Category IV suggests that substitution of wrong for correct letters is the form of error which is easiest to discriminate.

On balance the computer constructed forms functioned essentially equivalent to the WPC spelling test suggesting that this mode of test construction is worth further study. Separate error categories did not function as independently as was hoped. Greater care in balancing the content of tests as they are constructed and some experimentation with the definition of " super categories" could produce superior results.

References

Allen, Doris & Ager, Joel. "A Factor Analytic Study of the Ability to Spell". Educational and Psychological Measurement, 25, no. 1, 1965, 153-161.

Alper, Thelma G. "A Diagnostic Spelling Scale for the College Level, Its Construction and Use". Journal of Educational Psychology, 33, no. 4, 1942, 273-289.

Book, William F. & Harter, Richard S. "Mistakes which Pupils Make in Spelling". Journal of Educational Research, 19, 1929, 106-118.

Brody, D.S.A. "A Comparative Study of Different Forms of Spelling Tests". Journal of Educational Psychology, 1944, 35, 129-144.

Costello, C. G. "Anxiety, Intelligence and Discrimination". Journal of Abnormal Psychology, 1968, 73, no. 6, 594.

Foran, Thomas G. The Psychology and Teaching of Spelling. Washington, D.C: The Catholic Education Press, 1934.

Fremer, John & Anastasio, Ernest J. "Computer-Assisted Item Writing---I (Spelling Items)." Test Development Memorandum. Princeton, New Jersey: Educational Testing Service, July, 1968.

Furness, Edna L., Spelling for the Millions. New York: Appleton-Century, 1964.

Gates, Arthur I. A List of Spelling Difficulties in 3876 Words. New York: Teachers' College, Columbia University, 1937.

Hollingworth, Leta S. The Psychology of Specific Disability in Spelling. New York: Teachers' College, Columbia University, 1918.

Jackson, J. "The Influence of Machine-Scoring on Spelling Test Results". Journal of Psychology, 1943, 15, 277-83.

References continued

Knoell, Dorothy & Harris, C. W. "A Factor Analysis of Spelling Ability". Journal of Educational Research, 1952, 46, 95-111.

Lindquist, E. F. & Cook, W. W. "Experimental Procedures in Test Evaluation". Journal of Experimental Education, 1933, 1, 163-185.

Masters, Harry V. "A Study of Errors in Common Difficult Words". Studies in Education, 4, no. 4, Sept., 1927, 80.

Moore, J. E. "A Comparison of Four Types of Spelling Tests for Diagnostic Purposes". Journal of Experimental Education, 6, 1937, 24-28.

Nelson, M. J. & Denny, E. C. "Multiple-Choice Spelling Tests". School and Society, 44, 1936, 15-17.

Sturdyvin, E. M. "Note on Recognition versus Recall as Methods of Testing Spelling". Journal of Educational Psychology, 1937, 28, 394-6.

Thomas, Macklin. "A Machine-Scorable Test of Spelling Ability Based on Word Completions". Educational and Psychological Measurement, 28, 1968, 833-40.

Thorndike, Edward L. & Lorge, Irving. The Teachers' Word Book of 30,000 Words. New York: Bureau of Publication, Columbia University, 1944.

Thurstone, L. L. "Psychological Implications of Factor Analysis". American Psychologist, 3, 1948, 402-408.

Tidyman, W. F. & Johnson, Edith. "The Value of Grouping Words According to Similar Difficulties in Spelling". Journal of Educational Research, 10, 1925, 297-301.

Washington Pre-College (WPC) Testing Program. Counselor's Manual, 1968-1969. Seattle: WPC Testing Program, 1968.

References continued

Wesman, A. G. "The Usefulness of Correctly Spelled Words in a Spelling Test". Journal of Educational Psychology, 1946, 37, 242-6.

Weller, L. & Broom, M. E. "A Study of the Validity of Six Types of Spelling Tests". School and Society, 1934, 40, 103-4.

Woodson, M. I. "Computer Produces Forms with Randomly Selected and Ordered Items". Educational and Psychological Measurement, 1968, 28, no. 3, 857-8.